Math 164: Problem Set 5

Owen Jones

2/9/2024

8.1 Let $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and let $\mathbf{b} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$. $f(x_1, x_2) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{x}^\top \mathbf{b} + 3$

$\nabla f(x_1, x_2) = \mathbf{Q}\mathbf{x} + \mathbf{b}$

Thus, $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \frac{\nabla f^{(k)\top} \nabla f^{(k)}}{\nabla f^{(k)\top} \mathbf{Q} \nabla f^{(k)}} \nabla f^{(k)}$.

$x^{(1)} = [0, 0]^\top - \frac{\mathbf{b}^\top \mathbf{b}}{\mathbf{b}^\top \mathbf{Q}\mathbf{b}} \mathbf{b} = \frac{5}{6}\mathbf{b} = [-\frac{5}{6}, -\frac{5}{12}]^\top$

$\nabla f(x^{(1)}) = [\frac{1}{6}, -\frac{1}{3}]^\top$, $\frac{\nabla f^{(k)\top} \nabla f^{(k)}}{\nabla f^{(k)\top} \mathbf{Q} \nabla f^{(k)}} = \frac{\frac{5}{36}}{\frac{1}{4}} = \frac{5}{9}$.

$[-\frac{45}{54}, -\frac{45}{108}]^\top - [\frac{5}{54}, -\frac{20}{108}]^\top = [-\frac{25}{27}, -\frac{25}{108}]^\top$

Thus, $x^{(2)} = [-\frac{25}{27}, -\frac{25}{108}]^\top$.

We want to find an $\mathbf{x}$ that satisfies the FONC i.e $\nabla f = \mathbf{0}$. $\mathbf{Q}$ is non-singular, so we choose $\mathbf{x} = -\mathbf{Q}^{-1}\mathbf{b} = [-1, -\frac{1}{4}]^\top$.

8.4 (a) $\lim_{k\to\infty} 2^{-2^{k^2}} = 0$ because $\lim_{k\to\infty} -2^{k^2} = -\infty$.

(b) $\lim_{k\to\infty} \frac{\mathbf{x}^{(k+1)} - x^*}{(\mathbf{x}^{(k)} - x^*)^p} = \lim_{k\to\infty} \frac{2^{-2^{(k+1)^2}}}{2^{-p2^{k^2}}} = \lim_{k\to\infty} 2^{-2^{(k+1)^2} + p2^{k^2}}$ $= \lim_{k\to\infty} 2^{-2^{k^2+2k+1} + p2^{k^2}} = \lim_{k\to\infty} 2^{-2^{k^2}(2^{2k+1} - p)} = 0$ for any value of p because $\lim_{k\to\infty} -2^{k^2}(2^{2k+1} - p) = -\infty$ for any value of p. Thus, $\mathbf{x}^{(k)}$ has order of convergence ∞.

8.7 (a) By FONC, want $f'(x) = 0 \Rightarrow x^* = \frac{b}{a}$ minimizes f since $a > 0$.

(b) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha(a\mathbf{x}^{(k)} - b) = (1 - \alpha a)\mathbf{x}^{(k)} + \alpha b$.

(c) Assuming $\mathbf{x}^{(k)}$ converges, $x^* = (1 - \alpha a)x^* + \alpha b$
$\Rightarrow \alpha a x^* = \alpha b$
since $\alpha \neq 0, x^* = \frac{b}{a}$

(d) $\lim_{k\to\infty} \frac{|\mathbf{x}^{(k+1)} - x^*|}{|\mathbf{x}^{(k)} - x^*|^p} = \lim_{k\to\infty} \frac{|(1-\alpha a)\mathbf{x}^{(k)} + \alpha b - \frac{b}{a}|}{|\mathbf{x}^{(k)} - \frac{b}{a}|^p}$ $= \lim_{k\to\infty} \frac{|(1-\alpha a)\mathbf{x}^{(k)} - (1-\alpha a)\frac{b}{a}|}{|\mathbf{x}^{(k)} - \frac{b}{a}|^p} = \lim_{k\to\infty} |1 - \alpha a||\mathbf{x}^{(k)} - \frac{b}{a}|^{1-p}$. Because the limit only converges to a finite nonzero number for $p = 1$ and diverges to infinity for $p > 1$, the sequence converges linearly.

(e) The sequence converges iff $|1 - \alpha a| < 1$, so $0 < \alpha < \frac{2}{a}$.

8.8 We rewrite the function $f(x)$ using its quadratic form i.e $f(x) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} - \mathbf{x}^\top \mathbf{b} + \mathbf{c}$ where $\mathbf{Q} = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}$, $\mathbf{b} = [-5, -6]^\top$, and $\mathbf{c} = 7$.

Since $\mathbf{x}^{(k)}$ is a fixed step size algorithm, the sequence converges iff $0 < \alpha < \frac{2}{\lambda_{\max}\mathbf{Q}}$.
$\det(\lambda \mathbf{I}_2 - \mathbf{Q}) = \lambda^2 - 12\lambda + 20$ with roots located at $\lambda = 2, 10$, so the sequence converges for $0 < \alpha < \frac{1}{5}$.

8.13 Using the formula from lemma 8.1, we have $V(\mathbf{x}^{(k+1)}) = (1-\gamma_k)V(\mathbf{x}^{(k+1)})$ where $V(\mathbf{x}) = f(\mathbf{x})$ and $\gamma_k = 4\alpha \cdot 2^{-k}(1 - \alpha 2^{-k})$.
Since $0 < \gamma_k < 1$ for all k, we have $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$ so long $x^{(k)} \neq 1$ for $k \geq 0$, so the sequence satisfies the descent property.
$\displaystyle\sum_{k=0}^{\infty} \gamma_k = \sum_{k=0}^{\infty} 4\alpha 2^{-k} - \sum_{k=0}^{\infty} 4\alpha^2 2^{-2k} = 8\alpha - \frac{16}{3}\alpha^2 < \infty$, so the sequence is not globally convergent.

8.15 (a) $f(x) = (\mathbf{a}x - \mathbf{b})^\top (\mathbf{a}x - \mathbf{b}) = \mathbf{a}^\top \mathbf{a}x^2 - 2\mathbf{a}^\top \mathbf{b}x + \mathbf{b}^\top \mathbf{b} = \|\mathbf{a}\|^2 x^2 - 2\mathbf{a}^\top \mathbf{b}x + \|\mathbf{b}\|^2$
By the FONC, the only minimizer candidate is when $f'(x) = 0 \Rightarrow x^* = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|^2}$. Since $f''(x) = 2\|\mathbf{a}\|^2 > 0$, x^* is the minimizer for f.

(b) Since the sequence is a fixed step size gradient algorithm, $0 < \alpha < \frac{2}{\lambda_{\max}\mathbf{Q}}$ with $\mathbf{Q} = 2\|a\|^2$. Thus, $0 < \alpha < \frac{1}{\|\mathbf{a}\|^2}$.

8.16 (a) $f(x) = |\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = (\mathbf{A}\mathbf{x} - \mathbf{b})^\top (\mathbf{A}\mathbf{x} - \mathbf{b})$
$= (\mathbf{x}^\top \mathbf{A}^\top - \mathbf{b}^\top)(\mathbf{A}\mathbf{x} - \mathbf{b})$
$= \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{A}^\top \mathbf{b} - \mathbf{b}^\top \mathbf{A}\mathbf{x} + \|\mathbf{b}\|^2$
$= \mathbf{x}^\top (\mathbf{A}^\top \mathbf{A})\mathbf{x} - 2\mathbf{b}^\top \mathbf{A}\mathbf{x} + \|\mathbf{b}\|^2$
$= \mathbf{x}^\top (\mathbf{A}^\top \mathbf{A})\mathbf{x} - 2(\mathbf{A}^\top \mathbf{b})^\top \mathbf{x} + \|\mathbf{b}\|^2$
$\Rightarrow \nabla f(\mathbf{x}) = 2(\mathbf{A}^\top \mathbf{A})\mathbf{x} - 2(\mathbf{A}^\top \mathbf{b}), \mathbf{F}(x) = 2(\mathbf{A}^\top \mathbf{A})$

(b) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - 2\alpha \mathbf{A}^\top (\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b})$

(c) $2(\mathbf{A}^\top \mathbf{A}) = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \Rightarrow 0 < \alpha < \frac{2}{\lambda_{\max}(2(\mathbf{A}^\top \mathbf{A}))} = \frac{1}{4}$

8.18 The steepest descent algorithm is $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \frac{\mathbf{g}^{(k)\top}\mathbf{g}^{(k)}}{\mathbf{g}^{(k)\top}\mathbf{Q}\mathbf{g}^{(k)}}\mathbf{g}^{(k)}$ where $\mathbf{g} = \nabla f$.
If $\mathbf{x}^{(1)} = \mathbf{Q}^{-1}\mathbf{b}$ then $\mathbf{Q}^{-1}\mathbf{b} = \mathbf{x}^{(0)} - \alpha_0 \mathbf{g}^{(0)} \Rightarrow \mathbf{b} = \mathbf{Q}\mathbf{x}^{(0)} - \alpha_0 \mathbf{Q}\mathbf{g}^{(0)}$
$\Rightarrow \alpha_0 \mathbf{Q}\mathbf{g}^{(0)} = \mathbf{Q}\mathbf{x}^{(0)} - \mathbf{b}$.
Since $\mathbf{x}^{(0)} \neq \mathbf{Q}^{-1}\mathbf{b} \Rightarrow \mathbf{g}^{(0)} = \mathbf{Q}\mathbf{x}^{(0)} - \mathbf{b} \neq \mathbf{0}$.
Thus, $\mathbf{Q}\mathbf{g}^{(0)} = \frac{1}{\alpha_0}\mathbf{g}^{(0)}$, so $\mathbf{g}^{(0)}$ is an eigenvector of $\mathbf{Q}$.
If $\mathbf{g}^{(0)}$ is an eigenvector of $\mathbf{Q}$, then there exists $\lambda \in \mathbb{R}$ s.t $\mathbf{Q}\mathbf{g}^{(0)} = \lambda \mathbf{g}^{(0)}$.
It follows $\mathbf{Q}x^{(1)} = \mathbf{Q}(\mathbf{x}^{(0)} - \alpha_0 \mathbf{g}^{(0)}) = \mathbf{Q}\mathbf{x}^{(0)} - \alpha_0 \lambda \mathbf{g}^{(0)}$
$= \mathbf{Q}\mathbf{x}^{(0)} - \frac{\mathbf{g}^{(k)\top}\mathbf{g}^{(k)}}{\lambda \mathbf{g}^{(k)\top}\mathbf{g}^{(k)}}\lambda \mathbf{g}^{(k)} = \mathbf{Q}\mathbf{x}^{(0)} - \mathbf{g}^{(0)}$

Since $\mathbf{g}^{(0)} = \mathbf{Q}\mathbf{x}^{(0)} - \mathbf{b} \Rightarrow \mathbf{Q}x^{(1)} = \mathbf{b} \Rightarrow x^{(1)} = \mathbf{Q}^{-1}\mathbf{b}$, the sequence converges in one step.

8.21 (a) $f(x) = \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} - \mathbf{x}^\top\mathbf{b} + c$ where $\mathbf{Q} = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, c = 1$.
$0 < \alpha < \frac{2}{\lambda_{\max}\mathbf{Q}} = \frac{1}{5}$, so the largest step size possible is $\frac{1}{5}$.

(b) $f(x) = \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} - \mathbf{x}^\top\mathbf{b} + c$ where $\mathbf{Q} = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -16 \\ -23 \end{bmatrix}, c = \pi^2$.
$0 < \alpha < \frac{2}{\lambda_{\max}\mathbf{Q}} = \frac{1}{5}$, so the largest step size possible is $\frac{1}{5}$.

8.22 $\gamma_k = \beta\frac{\mathbf{g}^{(k)\top}\mathbf{g}^{(k)}}{\mathbf{g}^{(k)\top}\mathbf{Q}^{-1}\mathbf{g}^{(k)}}(2\frac{\mathbf{g}^{(k)\top}\mathbf{g}^{(k)}}{\mathbf{g}^{(k)\top}\mathbf{Q}\mathbf{g}^{(k)}} - \beta\frac{\mathbf{g}^{(k)\top}\mathbf{g}^{(k)}}{\mathbf{g}^{(k)\top}\mathbf{Q}\mathbf{g}^{(k)}})$
$= \beta(2-\beta)\frac{(\mathbf{g}^{(k)\top}\mathbf{g}^{(k)})^2}{(\mathbf{g}^{(k)\top}\mathbf{Q}^{-1}\mathbf{g}^{(k)})(\mathbf{g}^{(k)\top}\mathbf{Q}\mathbf{g}^{(k)})}$
$0 < \frac{\lambda_{\min}\mathbf{Q}}{\lambda_{\max}\mathbf{Q}} \leq \frac{(\mathbf{g}^{(k)\top}\mathbf{g}^{(k)})^2}{(\mathbf{g}^{(k)\top}\mathbf{Q}^{-1}\mathbf{g}^{(k)})(\mathbf{g}^{(k)\top}\mathbf{Q}\mathbf{g}^{(k)})}$ because $\mathbf{Q} > 0$.
If $0 < \beta < 2$, then $\beta(2-\beta) > 0$, so $\gamma_k \geq \beta(2-\beta)\frac{\lambda_{\min}\mathbf{Q}}{\lambda_{\max}\mathbf{Q}} > 0$. Because $\displaystyle\sum_{k=0}^{\infty}\gamma_k \geq \sum_{k=0}^{\infty}\beta(2-\beta)\frac{\lambda_{\min}\mathbf{Q}}{\lambda_{\max}\mathbf{Q}} = \infty$, the sequence converges globally to $\mathbf{x}^*$.
If $\beta \leq 0$ or $2 \leq \beta$, then $\beta(2-\beta) \leq 0$. It follows $\gamma_k \leq \beta(2-\beta)\frac{\lambda_{\min}\mathbf{Q}}{\lambda_{\max}\mathbf{Q}} \leq 0$.
Thus, $\displaystyle\sum_{k=0}^{\infty}\gamma_k$ does not diverge to ∞. Hence, if $\displaystyle\sum_{k=0}^{\infty}\gamma_k = \infty \Rightarrow 0 < \beta < 2$.